

REPORT DOCUMENTATION PAGEForm Approved
OMB NO. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE 31 May 2002		3. REPORT TYPE AND DATES COVERED Final, 10 Sep 1996 - 9 Mar 2002	
4. TITLE AND SUBTITLE A Proposal to Develop a Biotechnology Information Facility				5. FUNDING NUMBERS G: DAAH04-96-1-0145 0415	
6. AUTHOR(S) John Spalding, Sherri Clark, Peter Lammers					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) New Mexico State University Physical Science Laboratory P.O. Box 30002, Las Cruces, NM 88003-8002				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211				10. SPONSORING / MONITORING AGENCY REPORT NUMBER N/A 35919.1-LS-BIF	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
12 a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.				12 b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The objective of this grant was to develop facilities and information resources that support current research in biotechnology and to meet the goal of strengthening the biological science programs at HBCUs/MIs. As a result, the National Biotechnology Information Facility (NBIF) was established at New Mexico State University (NMSU). NBIF retained that name until September 2001, at which time its name was changed to the Southwest Biotechnology and Informatics Center (SWBIC). SWBIC was given a 6-month, no-cost extension, which changed the end of the period of performance to 9 March, 2002. SWBIC has become a recognized center for the biotechnology information resources presented through its website http://www.swbic.org/ . The most important accomplishments achieved during the grant are presented in this report, and include substantial work in education and outreach to promote biotechnology careers at HBCUs/MIs. Also, the capital improvements funded by the grant have resulted in a world-class biotechnology and bioinformatics facility in the third floor of the new Biochemistry wing at NMSU. It is the goal of SWBIC to maintain these resources and continue the goal of supporting biotechnology education and research both here and at other HBCUs/MIs.					
14. SUBJECT TERMS Biotechnology, bioinformatics, information, facility, education, outreach, minority				15. NUMBER OF PAGES 30 (numbered to 28)	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION ON THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED		20. LIMITATION OF ABSTRACT UL	

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. Z39-18
298-102

Enclosure 1

20030515 129

FINAL PROGRESS REPORT

A PROPOSAL TO DEVELOP A BIOTECHNOLOGY INFORMATION FACILITY

Authors: John B. Spalding, Sherri Clark, Peter J. Lammers

Performing Organization:

New Mexico State University
Physical Science Laboratory
P.O. Box 30002
Las Cruces, NM 88003

Sponsoring Agency:

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

Funding Number: DAAH04-96-1-0415

31 May 2002

Table of Contents

1	Statement of problem studied.....	1
2	Summary of most important accomplishments.....	1
2.1	Management and Sustainment.....	1
2.2	Capital Improvements.....	3
2.3	Informatics.....	4
2.4	Biotechnology.....	6
2.4.1	Core Wet-Lab and BSL3 Facilities	6
2.4.2	Science Projects.....	7
2.5	Education and Outreach.....	11
2.5.1	Training Activities for K-12 Teachers and Students.....	12
2.5.2	Training Activities for Post-Secondary Faculty, Staff and Graduate Students	16
2.5.3	Online K-12 Activities & Materials	19
2.5.4	Online Post Secondary Courses and Tutorials	20
2.5.5	Educational Games.....	21
2.5.6	Other Resources and Activities.....	22
3	Publications and technical reports.....	23
3.1	Papers published in peer-reviewed journals.....	23
3.2	Papers published in non-peer-reviewed journals or in conference proceedings	23
3.3	Papers presented at meetings, but not published in conference proceedings	24
3.4	Manuscripts submitted, but not published.....	25
3.5	Technical reports submitted to ARO	26
4	Participating Scientific Personnel	26
5	Report of Inventions	28
6	Bibliography	28
7	Appendixes.....	28

List of Appendixes, Illustrations and Tables

Table 1.	Description of SWBIC-funded science projects	8
Table 2.	Workshops for K-12 science educators and students	14
Table 3.	Post-Secondary Workshops	17

1 STATEMENT OF PROBLEM STUDIED

In the original Broad Agency Announcement DAAH04-96-R-BAA2 for a Biotechnology Information Facility, the objective was to develop facilities and information resources that support current research in biotechnology and to meet the goal of strengthening the biological science programs at HBCUs/MIs. Upon winning the award in September 1996, the National Biotechnology Information Facility (NBIF) was established at New Mexico State University (NMSU). NBIF retained that name until September 2001, the end of the five-year period of the grant. At that time, its name was changed (with approval of the COTR) to the Southwest Biotechnology and Informatics Center (SWBIC). Although "SWBIC" is used in this report as the name of the facility funded by the grant, work done prior to September 2001 was under the name "NBIF." SWBIC was given a 6-month, no-cost extension by ARO, which changed the end of the period of performance to 9 March, 2002.

SWBIC has become a recognized center for the biotechnology information resources presented through its website <http://www.swbic.org/>. The most important accomplishments achieved during the grant are presented in section 2, and include substantial work in education and outreach to promote biotechnology careers at HBCUs/MIs. Also, the capital improvements funded by the grant have resulted in a world-class biotechnology and bioinformatics facility on the third floor of the new Biochemistry wing at NMSU. It is the goal of SWBIC to maintain these resources and continue the goal of supporting biotechnology education and research both here and at other HBCUs/MIs.

2 SUMMARY OF MOST IMPORTANT ACCOMPLISHMENTS

This section is organized by the four main work areas in the "Revised Technical and Cost Plan for the Biotechnology Information Facility" (referred to as the "Revised Technical Plan," below), which was approved by ARO in 2000. A fifth work area, "Capital Improvements", is also included, and it applies primarily to accomplishments prior to 2000. This section summarizes the most important activities on the grant, and references certain pages on the SWBIC website at <http://www.swbic.org> for additional detail.

2.1 MANAGEMENT AND SUSTAINMENT

Management of the SWBIC grant has remained as proposed and approved in the "Revised Management Plan for the Biotechnology Information Facility" with the subsequent approval of Dr. John B. Spalding as the PI. The Leadership Team is composed of Dr. Peter Lammers (Biotechnology Team), Dr. John Spalding (Informatics Team), and Dr. Sherri Clark (Education Team).

SWBIC currently supports six professional staff and 5 students. Over the 5 ½ year period of the grant, SWBIC has employed a total of 23 professional staff (including 3 post-doctoral positions)

Final Progress Report on DAAH 04-96-1-0145

and 62 students (undergraduate and graduate). Many of the students earned degrees while working for SWBIC. In addition to staff and students, 11 faculty members from NMSU and other universities have participated in SWBIC scientific and educational activities. For a list of these individuals see <http://www.swbic.org/about/staff.php> and section 4. We have also continued to develop personal contacts with other HBCU/MI faculty for both workshop and collaboration purposes, as described in previous Interim Progress Reports and section 2.5 below.

A major goal of the SWBIC grant was to become self-sustaining by the end of the grant. Several external Advisory Board meetings were held to gather input from representatives from academia and the private sector, and many activities within the last year of the grant were directed towards the goal of self-sustainment. The Leadership Team decided that the most promising avenues for further funding support for SWBIC were genetic testing and grants for education and bioinformatics research.

In 2001 six proposals were submitted to NSF, NIH, and the Glaxo Wellcome Foundation. Of these, two have been awarded:

- NIH: Bioinformatics Education for Minority Students and Faculty (part of the Minority Access to Research Careers program); 5 years starting June 2002; value: \$462,400
- NIH: Functional Genomics of Uncharacterized Protein Sequences; bioinformatics research to continue the SWBIC-funded PCAPSS project; 3 years: June 2001-2004; value: \$268,101. This grant can be renewed for an additional 3 years.

These grants will continue the educational and research activities of SWBIC. Although four proposals were not given awards, three of these received favorable comments and will be considered for resubmission later this year.

Three other proposals (valued at approximately \$250,000 each) were submitted in 2002 as part of the Physical Science Laboratory's Phyto-Remediation Project with Picatinny Arsenal. All three would make use of and partially support the Molecular Biology/SWBIC Core Wet-Lab Facility. We are waiting on award information at this time, but the Army has already expressed serious interest in the first project.

- Phyto-remediation of Depleted Uranium and Arsenic: This is a study of the uptake of heavy metals from soil by arbuscular mycorrhizal fungi, which are plant symbionts.
- Rapid detection of Biological Warfare Agents: This is a project to develop and test technologies for the rapid detection of the presence of possible biological warfare agents in the field using ion mobility spectrometry methods; it is synergistic with the next project.
- Confirmatory Testing for Biological Warfare Agents: This project will develop and test methods for the genetic testing of samples for the detection of possible biological warfare

agents in a centralized laboratory facility. This work is complementary and synergistic with the previous project

In 2001, five proposals were submitted by other institutions that included collaboration with SWBIC. Three of these were funded (all are HBCUs):

- North Carolina A&T University: North Carolina A&T Interdisciplinary Graduate Teaching in Biotechnology, Genomics and Bioinformatics: This program was funded by the US Department of Education for four years (2001-2004) with Dr Mary Smith as the program director. SWBIC will be providing bioinformatics workshops.
- North Carolina A&T University: Agricultural Education to Address Emerging Needs of the Genomics Era: This proposal was funded for 2002-2004 by the USDA as part of its 1890 Institution's Capacity Building Grants program. Dr. Mulumebet Worku is the program director for this project. SWBIC is providing course building support, aid in the establishment of a bioinformatics learning facility, and ongoing advice and identification of collaborative opportunities.
- Alabama A&M University: Building a Bioinformatics Base: Preparing students for the New Millenium: This proposal was also funded for 2002-2004 by the USDA as part of its 1890 Institution's Capacity Building Grants program. Dr. Allan Zipf is the program director for this project. SWBIC will be providing bioinformatics workshops.

The other major avenue for continued funding is genetic testing. The SWBIC grant allowed us to build up the capabilities of the SWBIC and Molecular Biology Program Core Wet-Lab Facility to support grant educational activities and NMSU researchers in biotechnology. From this we have created the SWBIC Genetic Testing Laboratory (GTL). In collaboration with the NMSU Physical Science Laboratory and the Molecular Biology Program, the GTL is poised to receive accreditation from the American Association of Blood Banks as an accredited parentage testing laboratory using state-of-the-art DNA-based methods. PSL has invested its own business development funds for this effort. Once accredited, revenues from the GTL will be shared among SWBIC and the Molecular Biology Program to continue their activities, and to support biotechnology research and education for NMSU in general.

2.2 CAPITAL IMPROVEMENTS

The grant provided for capital improvements through refurbishment of the shell space on the third floor of the new wing of the Chemistry & Biochemistry building. Except for the Biosafety Level 3 (BSL3) suite, this work was completed by May 1998. This included offices, a computer laboratory, and a wet-bench laboratory with an adjoining equipment room. All rooms except the last were furnished and provided with complete electrical and communication utilities, including network connections.

The BSL3 suite capital improvement was funded jointly by the SWBIC grant (\$150,000) with matching funds from NMSU building and renovation funds, and contributions from the office of the Vice President for Research and the Deans of Arts & Sciences and Agriculture & Home Economics. The facility is near completion and is expected to be operational by late 2002. This facility will enhance the institutional capability to undertake research on infectious agents and to provide a bioterrorism response capability in the Southern New Mexico borderland region.

2.3 INFORMATICS

The Informatics Team (previously named Computational & Information Technologies) has been responsible for all computer-based activities, the development and maintenance of the SWBIC website, support of the BioScience and Educational areas, and bioinformatics research made available through services on the website. The SWBIC site has received eight awards (see <http://www.swbic.org/about/about.php>). Traffic to the SWBIC web site more than doubled in each of the last three years, from an average of 250,000 hits/month in 1999 to 540,000 in 2000, to 1,040,000 in the last six months of 2001 (due to a software failure, the first six months of statistics were lost). 2002 promises an even larger jump in visitors, as the average number of hits/month for January and February was 1,474,000.

The SWBIC Computer Laboratory contains 24 general use PCs that are used by SWBIC student staff and Molecular Biology faculty, staff, and students. These PCs are also used for local bioinformatics workshops and as a PC laboratory for certain Molecular Biology, Biochemistry, and Biology classes. There are also seven PCs used by professional staff and two high-end PCs for computer intensive graphics work for the website. This laboratory represents a valuable teaching resource, and will continue to be used by SWBIC and other departments as a computer laboratory for NMSU classes.

The laboratory also contains a 32-node, Beowulf cluster that was purchased on the grant (<http://www.swbic.org/about/darwin.php>). This cluster is the server for SWBIC bioinformatics website services and as a high-throughput computational resource for bioinformatics research, in particular the PCAPSS project (described in section 2.4.2).

The SWBIC website is designed to provide a central source for biotechnology information, education, and research. It is divided into three sections: Educational Resources (which are described in section 2.5), Internet Resources, and Products and Services. The Internet Resources section contains almost 4,500 annotated links to websites and pages in all fields of biotechnology, with special attention to bioinformatics, education, minorities, and the DoD. This is the single most complete such collection in the world. Products and Services contains SWBIC-developed clip-art for use in biotechnology education, the SWBIC Industry Register of biotechnology companies, and SWBIC bioinformatics services (described next).

In addition to developing and maintain the website, the Informatics Team is responsible for computational support of SWBIC-funded BioScience projects (described in section 2.4.2) and educational programs, resources, and games (described in section 2.5). Some of the projects were

focused on educational deliverables (such as courses and tutorials), while others on bioinformatics research. Products from these latter projects were subsequently developed into a suite of bioinformatics services available to the public, which are available at <http://www.swbic.org/products/bioinfo/bioinfo.php>. These services are:

- **PCAPSS (Protein Classification through the Assessment of Predicted Secondary Structure):** A new fold-recognition tool for helping identify hypothetical protein sequences. From a single query protein sequence, it builds a hidden Markov model of predicted secondary structure to search the PDB for proteins of similar structure.
- **BLAST Filter:** Builds a set of related sequences by running BLAST 2.2 on DNA or protein query sequence, filtering the matches through a set of rules, and returning the complete sequences of the BLAST matches that pass all rules. Includes user-defined rules and a graphical display of High Scoring Segment Pairs aligned with the query. Useful in building a sequence set for multiple alignment, phylogeny, and hidden Markov models.
- **Batch BLAST:** A service provided by SWBIC to the NMSU research community. It runs the Basic Local Alignment Search Tool (BLAST) for many query sequences. The results of these BLASTs may be sorted by the queries that resulted in the most significant matches to database sequences. Graphical alignment of High Scoring Segment Pairs is also provided, as well as Genbank links for matched subject sequences.
- **GraphAlign:** Performs a new type of graphical analysis of global pairwise alignments between a query sequence and multiple subject sequences. It provides new methods for analyzing and visualizing how a query (protein, DNA or RNA) aligns with a set of related sequences.
- **TransForm:** Translates data files in either Phylip or Nexus format to the other format. This service also can validate the format of Nexus data files.
- **SeqCheck:** Provides several utilities for formatting and analyzing a file of one or more sequences, including line formatting, alphabet checker, and composition analysis.
- **SeqMake:** Generates up to 10,000 random DNA, RNA, protein or user-defined sequences. Sequences can be generated from an input sequence by shuffling or generated from standard or user-defined compositional frequencies. Sequence lengths can also be randomized in several ways.
- **SeqEST:** Prepares an NCBI dbEST sequence submission file based on a user-supplied file of sequences. A time-saver when preparing dbEST submission files that also assures that all obligatory fields are provided.
- **EnVirusAT (Enteric Virus Sequence Analysis Tools):** Dedicated to the acquisition and analysis of viral genomes of enteric viruses. It provides the means for researchers and

students to rapidly collect citations or nucleic acid and amino acid sequence information for these viruses and analyze the information on selected web-based software to suit the researcher's needs.

- **EXTNUM (EXTended NUMeric):** A C++ class implementation that provides the user with a floating point number type that extends the dynamic range (i.e., the allowable range of the exponent) for a double type floating point number, while preserving the approximate 15 digit precision of the double as well. It is useful in bioinformatics software that works with large exponents.

2.4 BIOTECHNOLOGY

The Biotechnology Team (previously named the BioScience Team) is that part of SWBIC responsible for development of laboratory capabilities and selection and support of SWBIC-funded science projects.

2.4.1 Core Wet-Lab and BSL3 Facilities

As proposed in the Revised Technical Plan, the SWBIC/Molecular Biology Core wet-lab facility has been strengthened by the purchase of equipment on the grant, in particular the ABI Prism 3100 Sixteen Capillary Sequencing system and the ABI PRISM 7700 Quantitative PCR Sequence Detection System. Both instruments are installed and in use in the Molecular Biology Core Facility, which shares space with SWBIC on the third floor of the Chemistry and Biochemistry building. This facility is now capable of supporting state-of-the-art biotechnology research.

The Biosafety Level 3 (BSL3) suite is also located near the SWBIC area in the Chemistry and Biochemistry building. This facility was funded jointly by SWBIC (\$150,000) with matching funds from NMSU building and renovation funds, and contributions from the office of the Vice President for Research and the Deans of Arts & Sciences and Agriculture & Home Economics. It will be a focal point for the research activities of several scientists currently at NMSU, where it will provide a controlled area for the safe growth and biochemical analysis of viral pathogens. The facility is near completion and is expected to be operational in late 2002.

These Core facilities have been developed to foster sustainment activities on several fronts. The SWBIC Genetic Testing Laboratory (GTL) will use some of the equipment from the Core Wet-Lab initially for human DNA identity testing (also discussed in section 2.1). Proceeds from the GTL will be used (in part) to continue SWBIC activities and strengthen the research and educational capabilities of the Molecular Biology Program. The GTL has received support across the board from NMSU. It is intended to grow to eventually provide forensic and clinical DNA-based testing services.

Other sustainment activities (grant and fee for service) envisioned to make use of the Core Wet-Lab for SWBIC sustainment include the following:

- DNA sequencing on the ABI Prism 3100 Capillary Sequencing system
- Custom design and kinetic quantification of fluorescent PCR assay's using the ABI Prism 7700 Sequence Detection system
- Research with restricted pathogens (including biowarfare agents) using the BSL3
- Computational analysis of sequence data on the Beowulf computer cluster

The Core facilities are available for research and educational use by NMSU faculty. Numerous graduate students have taken advantage of the Core equipment to learn about and apply advanced methods in biotechnology. The facility has also strengthened NMSU's infrastructure in biotechnology, and with that its potential to win grants.

2.4.2 Science Projects

SWBIC has funded science projects to support its mission. The most significant has been the PathDB (Metabolic Pathways Database) project. This effort was started as a collaborative effort between SWBIC and the National Center for Genome Resources (NCGR) at Santa Fe. NCGR was a member of the original team proposed for SWBIC. This funding served as the base upon which NCGR invested much of its own funds in the project (more than matching SWBIC funding), resulting in a powerful tool for the biotechnology research community. NCGR received some additional funding from SWBIC in 2001 to improve research community access to PathDB, including an improved User Manual and a WWW version of PathDB (before it was a downloadable Java application). For more information see <http://www.ncgr.org/pathdb/>.

Table 1 contains information on the science projects, including the PIs, project descriptions, and final status.

Table 1. Description of SWBIC-funded science projects.

Project Name	Principal Investigator(s)	Description	Status
PathDB	National Center for Genome Resources (NCGR), Santa Fe, NM	Metabolic Pathways Database. This has been a joint effort between SWBIC and the National Center for Genome Resources (NCGR) since early in the grant. It was initially to be a database of plant secondary metabolic pathways (named "PlaSMA"), but has since been expanded to be more general, including pathways from all types of organisms.	PathDB has evolved into a powerful database for the storing and analysis of metabolic pathway and gene expression information. For a complete description see http://www.ncgr.org/pathdb/ . PathDB is available to the research community at the above website.
PCAPSS	Dr. Peter J. Lammers and Dr. John B. Spalding (NMSU)	Protein Classification through the Assessment of Predicted Secondary Structure. This is a bioinformatics method for identifying the functions of uncharacterized proteins based on sequence information.	PCAPSS has proved successful in identifying proteins based on predicted secondary structure alone. PCAPSS is a service on the SWBIC website, as well as being the source for several other SWBIC bioinformatics tools. It has produced three presentations at scientific meetings and is being funded by NIH for continued research (see section 2.1).
PEINA	Dr. Colleen B. Jonsson (NMSU)	Proteins and Enzymes that Interact with Nucleic Acids. This is a relational database of chemical and annotative information for nucleic acid binding proteins, which control gene expression.	The database was fully implemented, with a graphical user interface tool for data entry and a web-based query and output results system. Data for over 75 proteins were entered. Dr. Jonsson feels that PEINA needs additional data before access is given to the public, and she has assumed responsibility to secure new grant funds to continue population of the database building on the work funded by SWBIC.
EnVirusSAT	Dr. Kevin Oshima (NMSU)	Enteric Virus Sequence Analysis Tools. This interactive and instructional tool provides the means for researchers and students to rapidly collect citations for nucleic acid and amino acid sequence information for enteric viruses and analyze the information on selected web-based software to suit the researcher's needs.	EnVirusSAT is a service on the SWBIC website under Bioinformatics Tools.

Final Progress Report on DAAH 04-96-1-0145

Soil Bio-remediation Tutorial	Dr. N. Khandan (NMSU)	Soil Bio-remediation: A computer-based tutorial on bioprocesses in soil bioremediation and wastewater treatment.	The tutorial is available as a PowerPoint presentation on the SWBIC website under Courses & Tutorials.
CLAGO	Dr. Peter J. Lammers and Dr. John B. Spalding (NMSU)	Comparative Local Analysis of Gene Organization. The goal of the project was to build a tool that allows rapid presentation of the local information content around any gene of interest across all sequenced bacterial genomes.	Planning work was performed in 2000, which resulted in our conclusion that the needs for such an application were being met by other web-based services. Instead, as proposed in the 2000 Interim Progress Report, a page of such resources was developed that serve as a central point from which researchers may select the service(s) most appropriate for their needs. This page is the "Bioinformatics & Genomics/ Genomics & DNA Sequence Analysis/Comparative Genomics" category within the SWBIC Internet Resources section of the website.
Information Extraction	Dr. Jim Cowie, Computing Research Laboratory (CRL, NMSU)	Information Extraction from Biochemistry Research Articles. This preliminary effort was a study to transfer text extraction technology from defense-related applications to biotechnology. CRL worked with NCGR to focus on information important to metabolic pathway information, as additional support for the PathDB project.	A final report "Information Extraction from Biochemistry Research Papers" was prepared (available on request), along with the following deliverables: 1) Test data reverse engineered from the <i>Arabidopsis</i> data base and texts. 2) Lexicons of proteins and interaction words, 3) grammar and design for the prototype information extraction system. This software system was developed and used, but is not available through SWBIC at this time. The results from this project will be used as the basis for seeking additional grants in this research area.
TransForm	Dr. Brook Milligan, (Biology, NMSU) Drs. Enrico Pontelli, Gopal Gupta, and Desh Ramjan, (Computer Science, NMSU)	Translator Technology for Bioinformatics Interoperability. The objective of this project was to develop software technology that promotes interoperability between bioinformatics applications, specifically to translate between commonly used and complex data formats used to represent multiple alignment and phylogenetic information.	This project was completed, although not all planned data formats were implemented. TransForm is a service on the SWBIC website under Bioinformatics Tools. It translates data files in either Phylip or Nexus format to the other format, and also validates the format of Nexus data files. The PIs intend to use the work as a base on which to pursue further funding to continue the effort.

Final Progress Report on DAAH 04-96-1-0145

EXTNUM	Dr. Brook Milligan (NMSU)	EXTended NUMeric. A C++ class implementation that provides the user with a floating point number type that extends the dynamic range (i.e., the allowable range of the exponent) for a double type floating point number. It is useful in bioinformatics software that works with large exponents.	EXTNUM is a set of downloadable software on the SWBIC website under Bioinformatics Tools.
--------	------------------------------	---	---

2.5 EDUCATION AND OUTREACH

The Education and Outreach Team has been responsible for achieving the Army's goal to both strengthen biological science programs at minority institutions and to increase the number and broaden the expertise of minority scientists. Therefore, over the course of the grant, SWBIC not only provided information, resources and services to the biotechnology community, but also targeted information and training to minority students, faculty, and researchers.

During the grant period, SWBIC employed 18 minority students and 6 minority staff. Several of these students earned related degrees while working in the program (see section 4). Also, the web site and workshop advertisement mailings and personal contacts were used to attract the participation of faculty from minority serving institutions in visiting faculty positions. Dr. Narayanan Perumal from Florida State University worked during the summers of 1999 and 2000 to produce an online course in basic computational biology. Dr. R.L. Bernstein from San Francisco State University worked with SWBIC during his sabbatical year (2000-2001) to produce and present several bioinformatics workshops and an online course in more advanced bioinformatics. Dr. Arlene Antilla from Northern New Mexico Community College worked during the summer of 2000 to produce training materials for high school and college faculty, and to enhance the procedure manuals for the online games. And, in the summer of 2000, Mr. John Palmer of Gadsden High School developed six webquests (section 2.5.4) for use in high school classrooms.

Other outreach to minorities included a page on the SWBIC web site containing local NMSU minority resources at <http://www.swbic.org/links/minority.php>, and four categories of minority resources in the SWBIC resource database: general Minority Resources, Associations & Societies, Career Information, and Funding Sources. Also, a majority of the K-12 and post-secondary workshops provided by SWBIC were presented at minority serving institutions or schools, as detailed in sections 2.5.2 and 2.5.3.

SWBIC personnel developed collaborative contacts with other professionals in biotechnology, bioinformatics, and education that have been invaluable in completing the educational objectives and will continue to benefit the program. A selection of these contacts is listed below:

- Dr. Amal Abu-Shakra, Department of Biology, North Carolina Central University
- Dr. Richard L. Bernstein, Department of Biology, San Francisco State University
- Ms. Marie Haaland Borchert, Science Advisor Coordinator for the Las Cruces Science Education Alliance
- Dr. Stuart M. Brown, Department of Cell Biology and Research Computing Resources, NYU Medical Center
- Dr. Goldie Byrd, Department of Biology, North Carolina Central University
- Dr. Sharon Cosloy, Department Biology, The City College of New York (CUNY)

- Dr. Sarwan Dhir, Department of Agricultural Research, Fort Valley State University
- Dr. Chris Fields, President, HayFields Consulting, Inc.
- Dr. Greg Gibson, Departments of Genetics and Statistics, North Carolina State University
- Dr. Craig Liddell, VP of Informatics, Paradigm Genetics, Inc.
- Ms. Karin Matray, Director, The Las Cruces Teacher's Center
- Dr. Clifford Mintz, Chief Education Officer, BioInsights, LLC
- Dr. Spencer Muse, Department of and Statistics, North Carolina State University
- Dr. Doug Muller, President, Corridor, Inc.
- Dr. C.S. Prakash, Center for Plant Biotechnology Research, Tuskegee University
- Dr. Shawn Sendlinger, Department of Chemistry, North Carolina Central University
- Dr. Jeff Thorne, Department of Genetics, North Carolina State University
- Dr. Gary Washington, Department of Chemistry, The U.S. Military Academy, West Point
- Dr. Bruce Weir, Director of the Bioinformatics Research Center, North Carolina State University
- Dr. Mulumebet Worku, Department of Animal Science, North Carolina A & T State University
- Dr. Zhao-Bang Zeng, Department of Genetics, North Carolina State University
- Dr. Allan Zipf, Department of Plant and Soil Science, Alabama A&M University

Two of these contacts with faculty from minority institutions have lead to collaborations on grant proposals that have been funded (see section 2.1). In these funded projects, SWBIC personnel are providing training and consulting services to HBCUs as they build institutional and curricular infrastructure in bioinformatics.

2.5.1 Training Activities for K-12 Teachers and Students

Two science lesson kits were produced and one other was supplemented by SWBIC for use by K-12 educators and are available by contacting SWBIC. The first was a set of lessons on DNA and basic genetics for elementary students. This kit includes lesson plans, activities, illustrations, and a list of easily obtainable supplies for a two week biology session. The kit was produced in collaboration with a local education consulting firm, Corridor, Inc., and is targeted at 3rd, 4th or 5th grade depending on the background of the students. The kit has been evaluated and improved based on assessments from both beginning and very experienced elementary teachers.

A second lesson kit, Biotechnology Science for the 8th Grade, was produced by a collaboration between SWBIC scientists/educators, the local Science Education Alliance (SEA), and the local school district. A local middle school field-tested the kit. This series of biotechnology lessons and activities that involve four middle school content areas have been

Final Progress Report on DAAH 04-96-1-0145

incorporated into the eighth grade curriculum at Vista Middle School. The overall theme of the lessons was "to investigate." The project began with a meeting of shareholders in December, 2000 and progressed through development of a series of lessons and activities, a series of reviews, modifications, and pilot studies, to a curriculum that could be exported to other schools.

Pilot studies involved all four content areas and a technology lab in a true middle school concept of planning and coordination. Student interest and progress was documented through assignments, newsletters, digital pictures, video tape, and a CD. Personnel from SWBIC were involved in all phases of this project. The success of this project and the proposed adoption by the Las Cruces Public Schools may be attributed to the interest and support by the shareholders and teachers and administrators of Vista Middle School.

The third science kit effort was also a collaboration with the SEA, and was set in place to complete a set of bilingual backpack science kits for the local elementary schools. These backpack kits are interactive science lessons for kindergarten through fourth graders that are checked out by the students and completed at home with the help of parents/guardians. The 500 backpacks produced are now in use by local elementary schools.

Thirteen workshops were presented (Table 2) specifically for teachers in K-12 schools with three other workshops presented mainly for students. These workshops exceeded the original SWBIC goal of providing information and training for 135 K-12 science teachers over the life of the grant. Through them we directly helped over 150 educators and more than 50 students. By reaching this number of faculty, we can conservatively estimate that we have increased the number of students seeing new biotechnology related subject matter by 3750 per year. We have also increased educators' knowledge and skills so that many can further pursue the additions of biotechnology-related materials to their curricula.

Table 2. Workshops for K-12 science educators and students.

Title	Location	Date	Audience	Description
Internet Tools for K-12 Teachers: Teaching Using the Dog Genome Database	New Mexico State University, Las Cruces, NM	August 4-8, 1997	K-12 Science Teachers	Basic genetics and molecular genetics information and methods using the Dog Genome Initiative and Database as an example and resource for teaching. Participants were also trained in basic web page development.
DNA	Camino Real Middle School, Las Cruces, NM	February 12, 1998	8 th grade students	A short workshop covering DNA and its basic structure and function in cells. It included a hands-on experiment and other activities.
Learning to Use the Internet from Life on Mars!	Las Cruces Teacher's Center, Las Cruces, NM	March 24, 1998	K-12 Science Teachers	A short workshop to introduce K-12 educators to using the Internet as a research tool. The subject "Life on Mars" was used as a search term.
Learning to Use the Internet from Life on Mars!	New Mexico State University, Las Cruces, NM	April 2, 1998	K-12 Science Teachers	As above.
DNA	Alameda Elementary School, Las Cruces, NM	April 21, 1998	4 th grade students	A basic, short workshop introducing the concepts of DNA as the genetic material and the purposes and desired outcomes of genetic engineering.
A World of Microbes on the Internet	New Mexico State University, Las Cruces, NM	June 8-12, 1998	K-12 Science Teachers	A professional development workshop that provides scientific information on microbes as well as technical training on web navigation, evaluation of Internet resources, and the creation of web-delivered information pages.
A World of Microbes on the Internet	Lubbock Regional Services Center, Lubbock, TX	June 23-35, 1998	K-12 Science Teachers	As above.
DNA and the Crime Scene	New Mexico State University, Las Cruces, NM	July 29, 1999	K-12 science teachers	A short workshop review of DNA structure and function, DNA fingerprinting forensic science, a crime scene scenario lab experiment with agarose gel electrophoresis.
DNA and the Crime Scene	New Mexico State University, Las Cruces, NM	July 30, 1999	K-12 science teachers	As above.

Extreme Biotechnology in the Classroom	Ruidoso High School, Ruidoso, NM	December 10-11, 1999	K-12 Science Teachers	Two day workshop that presents information on extremophiles, extremozymes, and recombinant DNA technologies; information correlation with the state and national science standards and benchmarks; information links to related topics in textbooks; and how to use the Internet as a research tool. Labs and hands-on activities were included.
Bioinformatics I & II	Biotech Industry Organization meeting BIO2000, Boston, MA	March 26, 2000	High School AP and College Level Science Teachers	Two half-day workshop sessions that provides a basic overview of bioinformatics with a hands-on computer laboratory.
DNA and the Crime Scene	Tularosa High School, Tularosa, NM	May 3, 2000	K-12 Students and Science Teachers	A half-day workshop that involves a review of the structure and function of DNA, a description of DNA fingerprinting and how it is used in forensic science, a crime scene scenario, and a lab experiment to separate DNA fragments by agarose gel electrophoresis.
Bioinformatics	Tufts University, Boston, MA	June 23, 2000	High School AP Science Teachers	A two hour workshop that introduces the concepts of bioinformatics and the SWBIC educational game "Origin: Unknown" as teaching tools for AP high school classrooms.
Introductory Genetics Workshop	Northern New Mexico Community College, Espanola, NM	September 16, 2000	K-12 Science Teachers	One day basic workshop on DNA, genetics and biotechnology.
Bioinformatics for Beginners	Biotech Industry Organization meeting BIO2001 Teacher Professional Development Program; San Diego, CA	June 24, 2001	High School AP and College Level Science Teachers	A half-day workshop introducing the concept of bioinformatics and related Internet-accessible tools.
Biotechnology in the Middle School Classroom	Las Cruces Teacher's Center, Las Cruces, NM	August 1-3, 2001	Middle School Science Teachers	Overview and training workshop to prepare middle school teachers to present the SWBIC developed 8 th grade biotechnology kit in the classroom

2.5.2 Training Activities for Post-Secondary Faculty, Staff and Graduate Students

SWBIC also sponsored activities for post-secondary faculty, staff and graduate students. These activities took the form of financial support for two of the NMSU Zia Symposia, support and instruction for an NMSU course in bioinformatics for undergraduate and graduate students, and twelve workshops that were developed and presented by SWBIC. These workshops took place across the country, mainly at NMSU and minority institutions in the southeastern and eastern US. These post-secondary workshops reached the original proposed goal of reaching approximately 135 faculty the great majority of which were from minority institutions. These workshops also directly reached a significant number of students who were either training for research positions or science teachers-in-training.

Table 3 shows the post-secondary workshops that were presented by SWBIC. The bioinformatics workshop presented at The City College campus of the City University of New York in 2001 was used to produce a video workshop on CD-ROM. This workshop is now available on CD-ROM and is anticipated to be available on the web site after some improvements are made to the audio quality.

Table 3. Post-Secondary Workshops.

Title	Location	Dates	Audience	Description
Nucleic Acid and Protein Sequence Analysis Workshop	New Mexico State University, Las Cruces, NM	March 30-31, 1997	Faculty, staff and graduate students	A basic introduction to sequence analysis concepts and tools.
Human Genome Database: Impact on the Prevalence of Diabetes	New Mexico State University, Las Cruces, NM	August 11-15, 1997	NMSU faculty, staff and students	A workshop on the Human Genome Project with specific reference to Type II diabetes.
DNA and the Crime Scene	New Mexico State University, Las Cruces, NM	October 26, 1998	NMSU science education students	A short workshop review of DNA structure and function, DNA fingerprinting forensic science, a crime scene scenario lab experiment with agarose gel electrophoresis.
Bioinformatics Tools on the Internet	Tuskegee University, Tuskegee, AL	November 20-21, 1998	Faculty, staff and graduate students	An overview of the bioinformatics / computational molecular biology tools available for use over the Internet
Bioinformatics: Merging Biology and Computer Science	North Carolina Central University, Durham, NC	October 21-23, 1999	Faculty, staff and graduate students	Overview of bioinformatics, bioinformatics tools and analysis techniques and advances in molecular biology research techniques.
DNA and the Crime Scene	New Mexico State University, Las Cruces, NM	November 30, 1999	NMSU science education students	A short workshop review of DNA structure and function, DNA fingerprinting forensic science, a crime scene scenario lab experiment with agarose gel electrophoresis.
Bioinformatics I & II	Biotech Industry Organization meeting BIO2000, Boston, MA	March 26, 2000	High school AP and college level science teachers	Two half-day workshop sessions that provide a basic overview of bioinformatics with a hands-on computer laboratory.
Bioinformatics	New Mexico State University, Las Cruces, NM	September 22-23, 2000	Faculty, staff and graduate students	Basic overview of bioinformatics and common publicly available sequence analysis tools.
Bioinformatics: Merging Biology and Computer Science	North Carolina A&T State University in Greensboro, NC	October 16-18, 2000	Faculty, staff and graduate students	Overview of bioinformatics, bioinformatics tools and analysis techniques, and advances in molecular biology research techniques.
Bioinformatics & Advanced Molecular Biology Techniques	The City College of the City University of New York, New York, NY	January 3-5, 2001	Faculty, staff and graduate students	As above.

Bioinformatics & Advanced Molecular Biology Techniques	Fort Valley State University, Fort Valley, GA	April 19-21, 2001	Faculty, staff and graduate students	As above.
Bioinformatics for Beginners	Biotech Industry Organization meeting BIO2001 Teacher Professional Development Program; San Diego, CA	June 24, 2001	High School AP and College Level Science Teachers	A half-day workshop introducing the concept of bioinformatics and related Internet-accessible tools.
Introductory Genomics & Computational Biology	New Mexico State University, Las Cruces, NM	November 10, 2001	Graduate students	An overview of Genomics and related WWW tools and demos with hands-on computer labs for the participants.

2.5.3 Online K-12 Activities & Materials

SWBIC has developed a number of specialized classroom activities designed to supplement K-12 science courses. These are available at <http://www.swbic.org/education/activites.php>. These include classroom activities and materials and laboratory activities and webquests, which are interactive computer based activities intended to guide students through using the Internet as a research tool. The following is a summary:

Class Activities & Materials

- ABO Blood Group - A classroom activity intended to familiarize students with human red blood cells, antibodies, and blood types.
- Transcription and Translation Visual Exercises - Quiz-type exercises that provide hands-on experience for students to visualize and fully understand the importance of replication, transcription, and translation.

Laboratory Activities

- DNA Fingerprinting: Analysis of Crime Scene DNA - This exercise examines DNA fingerprinting techniques used for identification, to compare the DNA of a crime suspect to those samples found at the crime scene.
- Enzymes and Laundry Detergent - A laboratory activity which allows students to investigate the effectiveness of enzymes in laundry detergents.
- Watch an Enzyme Work - A hands-on activity examining protease enzymes as they catalyze the hydrolysis of proteins into amino acids.
- Wheat Germ DNA Isolation Experiment - This activity answers the following questions: where is DNA located, how can DNA be isolated, can DNA be seen with the naked eye, and what does DNA look like?

Webquests

- Atoms, Ions, and Isotopes - A webquest and short review of atoms, ions, and isotopes.
- Biome WebQuest - The purpose of this webquest is to explore the characteristics of biomes.
- Blood-Brain Barrier Outline - This webquest is designed to introduce the phenomenon of the blood-brain barrier.

- Enzyme Unit Plan - A webquest designed to provide an understanding of the role of enzymes.
- Exploring Cellular Respiration - A Socratic webquest exploring the process of cellular respiration in photosynthetic organisms.
- Exploring Photosynthesis - A Socratic webquest which explores photosynthesis in organisms and the environment.

2.5.4 Online Post Secondary Courses and Tutorials

SWBIC has developed a number of specialized materials and tutorials designed to supplement biotechnology courses at NMSU and other universities and colleges. These are available at <http://www.swbic.org/education/courses.php>. The following is a summary:

Bioinformatics: Many detailed PowerPoint presentations suitable for introductory upper division or graduate courses in bioinformatics. The presentations are organized for lecture or reference, with an index describing their use. The course materials emphasize bioinformatics applied to DNA and protein sequences and protein structures. The analysis tools described are accessible through freely available Internet resources. Other topics like molecular phylogenetics and PCR primer design are also introduced, and some of the materials would also be suitable for use in genetics or genomics courses. These materials were developed by Dr. R. L. Bernstein of San Francisco State University while on sabbatical at SWBIC and will be presented in his bioinformatics course at SFSU. SWBIC resources and presentation materials have been incorporated into the slides.

Bioinformatics & Genome Analysis: This course has become Molecular Biology 470 at NMSU and was taught in the Spring of 2000 and 2001; it will continue to be taught each Fall semester. It introduces the basic concepts of the discipline of bioinformatics and introduces basic computational tools and methods that are used in the organizing and analysis of data. It includes an introduction to large-scale DNA sequencing methods, and the methods and algorithms used to perform analyses such as alignments, homology searches, and advanced searching techniques. In addition the course presents basic computational tools such as use of the WWW, Unix fundamentals, file management, and database concepts.

Computational Management of Molecular Biological Information: This basic, on-line, bioinformatics course was authored by Dr. Narayanan B. Perumal from Florida State University for the SWBIC web site. The course is presented in lecture notes format with accompanying laboratory exercises and provides a good overview of the fundamentals of Computational Molecular Biology. Links to a large number of bioinformatics related resources on the Internet are provided within the course modules.

Genome Sequencing, Fragment Assembly, and the use of Consed, Phred, and Phrap: A PowerPoint presentation on genome sequencing methods with a focus on how sequence fragments are assembled. Topics include: review of DNA structure, polymerase chain reaction, sequencing chemistry, raw sequence data, shotgun sequencing approach, and treatment and assembly of sequence fragments. The methods used by Phred and Phrap for base calling, sequence quality determination, and fragment assembly are treated extensively, including computer science techniques used by these programs.

Soil Bioremediation: A PowerPoint presentation giving reviews and case histories of four major bioremediation processes: bioventing, land treatment, slurry phase treatment, and composting.

Tutorial in Molecular Biology: A set of tutorials on retrieving a gene sequence or a family of gene sequences, constructing a family of aligned protein sequences, and displaying the 3-dimensional structure of a protein.

Introduction to Environmental Engineering: A course covering principles in environmental engineering and science: physical chemical systems and biological processes as applied to pollution control.

2.5.5 Educational Games

SWBIC has developed three complex, graphical, interactive, WWW games that have proven to be popular in schools across the nation (based on email responses and requests for information from students and teachers). These games have been presented to teachers at all SWBIC presented workshops subsequent to their release as well as at the National Science Teacher's Association Convention 1999, the Biotechnology Industry Organization Convention in 2000 & 2001, and at a teacher's professional development symposium at Tufts University in 2000. The games are available at <http://www.swbic.org/education/games.php>.

iDNAfication: Intended to teach advanced high school students to use biology and molecular biology techniques to solve forensics mysteries.

Origin: Unknown: Intended to teach students and science educators how to utilize WWW-based bioinformatics tools in the analysis of biological macro-molecular sequence data.

Outbreak!: Intended to aid in teaching advanced high school and college students clinical and diagnostic methodologies, and skills used in microbiology.

Two other games were also developed. Unlike the above, these are not WWW games, but are instructions and materials for playing the games in the classroom.

PCR Dash: Illustrates the Polymerase Chain Reaction (PCR) process for students.

Protein Explorations: An interactive game/tutorial to allow students to explore protein structure files in the public Protein Data Bank.

2.5.6 Other Resources and Activities

Several listserves were set up and facilitated by SWBIC staff and associated faculty to serve the NMSU and biotechnology communities. These are listed below. The subject matter covered by each is indicated by its name.

- biotech-talk
- invertebrate
- plantmetab
- toxicology-digest
- vertebrate
- watervirus

During the grant period SWBIC employees attended conference, training courses, and received relevant company training on SWBIC-owned laboratory equipment in an effort to support their continuing education.

- Dr. Clark attended the New England Biolabs Molecular Biology and PCR Summer Workshop in July 1999 at Smith College
- Dr. Clark attended an Advanced bioinformatics training course in June 2000 at MIT
- Dr. Lammers and Dr. Spalding attended a one-day tutorial of the prediction of 3-D protein structures at the ISMB conference in 2000.
- GTL staff were trained by SWBIC personnel on instrument usage.

The following meetings were attended by SWBIC scientific staff, with papers, posters, or workshops presented in the majority (see sections 2.5.2, 2.5.3, 3.2, and 3.3):

- Conference in In Vitro Biology, 1997
- 9th International Conference of the American Society for Horticulture Science, 1997
- 2nd Annual Conference on Computational Genomics, 1998
- 10th International Genome Sequencing and Analysis Conference (GSAC), 1998
- Second Georgia Tech International Conference on Bioinformatics, 1999
- Plant & Animal Genome VII Conference, 1999
- National Science Teachers of America Annual Meeting, 1999

- Association of Environmental Engineering and Science Professors, 1999
- 1st European Seminar on Environmental Engineering Education, Zurich, 1999
- Biotechnology Industry Organization Annual Meetings: BIO '99, BIO 2000, and BIO 2001
- Plant and Animal Genome-VIII Conference, 2000
- 9th International BioThermoKinetics Meeting, 2000
- Eighth International Conference on Intelligent Systems for Molecular Biology (ISMB 2000)
- 12th International Genome Sequencing and Analysis Conference (GSAC), 2000
- Association for Molecular Pathology Annual Meeting, 2000
- Cambridge Healthtech Institute's Fourth Annual Functional Genomics, 2000
- Intelligent Systems in Molecular Biology, 2001
- Plant and Animal Genome-IX Conference, 2001
- Symposium on Computation Cell Biology, 2001
- Gordon Conference on Bioinformatics, 2001
- 1st Virtual Bioinformatics and Genomics Conference, 2001
- Cambridge Healthtech Institute conference on Metabolic Profiling, 2001
- International Systems Biology Conference, 2001
- Gene Ontology Meeting, 2001

3 PUBLICATIONS AND TECHNICAL REPORTS

3.1 PAPERS PUBLISHED IN PEER-REVIEWED JOURNALS

- Rimer, M. and M. O'Connell. 1998. BioABACUS: a database of abbreviations and acronyms in biotechnology and computer science. *Bioinformatics* 14: 888-889.
- Iglesias, J.R. , Pontelli, E., Ranjan, D., Milligan, B., and Gupta, G. 2001. Interoperability between Bioinformatics Tools: A Logic Programming Approach. In *Practical Aspects of Declarative Languages, Lecture Notes in Computer Science*, Springer Verlag.
- Gupta, G. Pontelli, E., Ranjan D., Karshmer A., Milligan, B., and Guo, H 2002. Semantic Filtering: a Killer Application for Logic Programming. In *Practical Aspects of Declarative Languages, Lecture Notes in Computer Science*, Springer Verlag, 2002.
- Kuffner RM, Gonzales M, Steadman P, Wlodek DK, Jankowitz RJ, Boinoff JR, Montoya L, Peterson TF, Bulmore DL, Blanchard JL The Molecular Biology Database Collection: 2002 update - PathDB, *Nucleic Acids Research Online*.
<http://www3.oup.co.uk/nar/database/summary/115>

3.2 PAPERS PUBLISHED IN NON-PEER-REVIEWED JOURNALS OR IN CONFERENCE

PROCEEDINGS

- Lammers, P. J., Spalding, J. B., Duran, S. T. and Garcia, N.. 1998. Orphan Protein identification through comparisons of predicted versus experimental secondary structures. *Microbial & Comparative Genomics* 3, C-76-C-77 (abstract). Poster presented at the 10th International Genome Sequencing and Analysis Conference, 17-20 September, Miami Beach, FL.
- Mendes P, Bulmore D, Farmer A, Steadman P, Waugh M, Wlodek S. PathDB: A Second Generation Metabolic Database. In *Proceedings of the 9th International BioThermoKinetics Meeting*, 3-8 April 2000, pp207-212.
<http://www.sun.ac.za/biochem/btk/book/Mendes.pdf>

3.3 PAPERS PRESENTED AT MEETINGS, BUT NOT PUBLISHED IN CONFERENCE PROCEEDINGS

- Phillips, G., M. Mendoza, M. O'Connell, D. Bustamante, and E. Burlbaw. 1997. Single-point access to biotechnology information and analysis tools via the World-Wide Web. Conference on In Vitro Biology, Washington, DC, 14-18 June.
- Phillips, G., M. Mendoza, D. Bustamante, and E. Burlbaw. 1997. Web-based single-point access to biotechnology information and analysis tools. 9th Annual International Conference of the American Society for Horticulture Science, Salt Lake City, UT, 22-26 July.
- Lammers, Peter J., J. B. Spalding, and S. T. Duran. 1998. Using Hidden Markov models of predicted protein secondary structure states for gene identification and fold recognition. Poster presented at the 2nd Annual Conference on Computational Genomics, 1-3 November, Reston, VA.
- Lammers, Peter J., J. B. Spalding, and S. P. Duran. 1999. Use of secondary structure state hidden Markov models for gene identification and protein fold recognition. Poster presented at the Second Georgia Tech International Conference on Bioinformatics, 11-14 November, Atlanta, GA.
- Steadman, Peter, D. Bulmore, J. Bingham, A. Farmer, M. Waugh, P. Mendes, T. Wei, M. Rimer, J. Spalding, and B. Sobral. 1999. PLASMA: The Plant Secondary Metabolic Archive. Presentation at the Plant & Animal Genome VII Conference, 17-21 January, San Diego, CA.
- Clark, S. D. 1999. Internet-based Interactive Tools: Outbreak! Presentation to the National Science Teacher's Association 47th National Convention, March 25-28, Boston, MA.
- Khandan, N. N. 1999. Development of computer-based interactive tutorials and quizzes. Computer poster presentation at the Association of Environmental Engineering and Science Professors, College Park, PA, July 31-Aug 1.
- Khandan, N. N. 1999. Use of computer simulation models in distance education. Paper presentation at the 1st European Seminar on Environmental Engineering Education, Zurich, Switzerland, August 22-24.
- Waugh, M, Bulmore D, Farmer A, Steadman P, Wlodek S, Mendes P. PathDB: A Metabolic Database with Sophisticated Search and Visualization Tools. *Plant and Animal Genome-VIII*, January 9-12, 2000. Poster and Software Demonstration.

- Kuffner RK, Blanchard JL. Towards the Virtual Organism: Networks and Expression. Intelligent Systems in Molecular Biology. July 21-15, 2001 Three hour tutorial.
- Gonzales M, Waugh, M, Bulmore D, Farmer A, Steadman P, Wlodek S, Mendes P, Blanchard J. PathDB: A Second Generation Metabolic Database. Intelligent Systems in Molecular Biology, August 19-23, 2001. Presentation.
- Gonzales M, Kuffner RM, Steadman P, Wlodek DK, Jankowitz RJ, Boinoff JR, Montoya L, Peterson TF, Bulmore DL, Blanchard JL. PathDB: Making Metabolic Connections in *Arabidopsis*, Plant and Animal Genome-IX, January 13-17, 2001. Poster.
- Blanchard JL. A System for Building, Visualizing and Comparing Cellular Networks. Symposium on Computation Cell Biology, March 4-6, 2001. Poster.
- Kuffner RM, Gonzales M, Steadman P, Wlodek DK, Jankowitz RJ, Boinoff JR, Montoya L, Peterson TF, Bulmore DL, Blanchard JL. Analysis of Sequence and Expression Data in the Context of Biological Interactions, Intelligent Systems in Molecular Biology. July 21-15, 2001. Poster and Software Demo.
- Kuffner RK. Creating Model Cells from Molecular Interaction Data. Gordon Conference on Bioinformatics, Aug 19-24, 2001. Poster.
- Blanchard JL. Creating Model Cells from Molecular Interaction Data. 1st Virtual Bioinformatics and Genomics Conference, Oct 15-15, 2001. Invited speaker.
- Blanchard JL. A Platform for the Interpretation of Gene, Protein, and Metabolite Expression Data in the Context of Cellular Interactions. Cambridge Healthtech Institute Conference on Metabolic Profiling, Dec 3-4, 2001. Invited speaker.
- Blanchard JL. Analysis of Sequence and Expression Data in the Context of Biological Interactions. International Systems Biology Conference, Nov. 4-7, 2001. Two hour tutorial.
- Gonzales M, Kuffner RM, Steadman P, Wlodek DK, Jankowitz RJ, Boinoff JR, Montoya L, Peterson TF, Bulmore DL, Blanchard JL. A Tool for Curating Interaction Data. Gene Ontology Meeting, October, 2001. Presentation.

3.4 MANUSCRIPTS SUBMITTED, BUT NOT PUBLISHED

- Spalding, J. B., Duran, S. P. and Lammers, P. J. 1999. Hypothetical protein classification using hidden Markov models of predicted secondary structure. Submitted to Journal of Computational Biology. Not accepted.
- Spalding, J. B., C. John, and P. J. Lammers. 2001. BLAST Filter: rule-based identification and retrieval of sets of related protein and DNA sequences. Submitted to Bioinformatics. Not accepted.
- Clark, S. D. and Smith, G. B. 2001. Outbreak! – Teaching clinical and diagnostic microbiology methodologies with an interactive online game. Submitted to Journal of College Science Teaching. Under review.
- Spalding, J. B., C. John, and P. J. Lammers. 2002. BLAST Filter: rule-based identification and retrieval of sets of related protein and DNA sequences. To be submitted shortly to Genome Research.

3.5 TECHNICAL REPORTS SUBMITTED TO ARO

Interim Progress Report on DAAH 04-96-1-0145 for period 10-30 September, 1996
Interim Progress Report on DAAH 04-96-1-0145 for period January 1–December 31, 1997
Interim Progress Report on DAAH 04-96-1-0145 for period January 1–December 31, 1998
Interim Progress Report on DAAH 04-96-1-0145 for period January 1–December 31, 1999
Interim Progress Report on DAAH 04-96-1-0145 for period January 1–December 31, 2000

4 PARTICIPATING SCIENTIFIC PERSONNEL

A large number of scientific personnel were employed under the SWBIC grant through its 5 ½ year period. Those earning advanced degrees during this time are so noted. Personnel are divided among the following categories and are faculty, staff, or students of NMSU unless otherwise indicated:

Leadership:

John B. Spalding, PhD, PI and Leader of Informatics Team
Peter J. Lammers, PhD, Leader of Biotechnology Team
Sherri Clark, PhD, Leader of Education Team

Current Professional Staff:

Stefan Long, Scientist (2001-)
Gabriel Rivera, Scientist (1998-), earned BA in Biology
Patrick McDougle, Software Engineer/Programmer (1997-), earned BS in computer science

Current Students:

Amanda Bird (2001-)
Audrey Franco (2000 -), earned BS in communication disorders
Oliver Hampton (2001-)
Ciju John (2000-), earned BS in Computer Science
Kelly Marshall (2000-), earned BS in Microbiology

Participating NMSU Faculty:

James Botsford, PhD, Department of Biology
Fernando Cadena, PhD, PE, Civil, Agricultural & Geological Engineering
Jim Cowie, PhD, Computing Research Laboratory
Gopal Gupta, PhD, Department of Computer Science
Colleen Jonsson, PhD, Department of Chemistry and Biochemistry
N. Khandan, PhD, PE, Civil, Agricultural & Geological Engineering
Brook Milligan, PhD, Department of Biology
Mary O'Connell, Ph.D, Agronomy and Horticulture
Kevin Oshima, PhD, Department of Biology
Greg Phillips, Ph.D, Agronomy and Horticulture
Enrico Pontelli, PhD, Department of Computer Science
Desh Ranjan, PhD, Department of Computer Science

Visiting Faculty:

Richard Bernstein PhD, Department of Biology, San Francisco State University, San Francisco, CA

Final Progress Report on DAAH 04-96-1-0145

Stuart Brown, PhD, New York University - Medical School, New York, NY (co-taught bioinformatics workshops)

Narayanan Perumal, PhD, Florida State University, Tallahassee, FL

SWBIC Alumni, Professional Staff:

Arlene Antilla PhD, Scientist (2000-2001)
Edward Burlbaw PhD, Scientist (1996-2000)
Donald Bustamante, Scientist (1996-2001)
Cristina Chacon, Physical Scientist (1997)
Janet Greenlee, Physical Scientist (1998)
Reid Hayhow, Scientist (1998-2000)
Stephen B. Hottman, PI (2000-2001)
Chris Hoxworth, Scientist (1997-2001)
Bernie Margolis, Scientist (1999-2000)
Sean O'Brien PhD, Scientist (1998-2001)
Mendell Rimer, PhD, Post-doctoral scientist (1997-1998)
Alex J. Ryncarz II, Ph.D., Scientist (2000-2002)
Mike Strand, Scientist (1997-1999)
Selene Virk, Scientist (1999-2000)
Tao Wei, PhD, Post-doctoral scientist (1997-1998)

SWBIC Alumni, Students, showing degrees earned:

Sarah Anderson (1999-2000), earned BS in biology
Aaron Begay (1997)
AK Beyer (1997)
Mgavi Brathwaite, Student Intern (1999) from Tuskegee University
Sal Briones (1997)
Carlos Cantu (2000)
Ed Carpenter (1998)
Guoyou Chen (2000)
Praveen Cherukuri (2000-2001), earned MS in Biochemistry
Hae-Ri Choe (2001)
Padmaja Chowti (2001)
Samantha Cicero (1998), earned MS in Biology
Daniel Cloyd (1999-2000)
Roy Coble (1997)
Tari Cook (1997)
Jeff Dage (1998)
Bo Du (2000-2001)
Zhidian Du (2001)
William English (1999)
Michael Fajardo (1997)
David Fashena (1997)
Ephraim Ford (1998-2001), earned BS in Engineering
Antonio Fortes (1997)
Jennifer Gonzales (1997)
Marco Gonzalez (2000-2001)
Vishnu Govindarajulu (1998-1999)
Laura Haas (1998), earned MS in Biology
Joe Horton (1997), earned MS in Computer Science

Final Progress Report on DAAH 04-96-1-0145

Juan Iglesias (2000-2001), earned PhD in Computer Science
Anuradha Jayanti (2001), earned MS in Computer Science
Mary Jones (2000-2001)
Daniel Kim (2000)
Gem Kosan (1999-2000), earned BS in Computer Science
Chaitanya Kudikala (2000-2001)
Robert Lamb (1997)
Gil Lara (1999)
Chongbing Liu (2000-2001)
Ravindra Matur (1997)
Erik Measure (2001), high school student, earned diploma
Joseph T. Meier (1997)
Marcus Mendoza (1996-1997)
Shahram Mori (1998), earned PhD in Molecular Biology
Sanjeev Nirmalakhandan (1998-2001), earned BS in Computer Science
Pramod Pesara (2000-2001)
Harish Ramanathan (2001)
Tobias Rodriguez (1997)
Jason Rosenberg (1997)
John Schutte (1997)
Phillip Seitsinger (1998), earned BS in Biology
Joshua Shagam (2001)
David Sultemeier (1999), earned BS in Biology 2001
Jennifer Tabullo (1999-2000), earned BS in Engineering
Pradeep Thota (2000-2001)
Michael Wildman (1999-2000)
Teresa Wildman (1998-2000), earned BS in Computer Science
Liansheng Zheng (1997)

5 REPORT OF INVENTIONS

Nothing was invented and no patents were applied for during the period of performance.
Software under development may be copyrighted at a later date.

6 BIBLIOGRAPHY

N/A

7 APPENDIXES

N/A